COMPUTER INPUT MICROFILM (CIM) FEASIBILITY STUDY

ARS-NE-46
July 1974

AGRICULTURAL RESEARCH SERVICE • U.S. DEPARTMENT

ACKNOWLEDGMENTS

The authors gratefully acknowledge the cooperation and assistance of the following individuals and organizations in this study. Clarence G. Beale, Jr., manager of the Scientific Applications Project System, Application Division, National Oceanic and Atmospheric Administration, Suitland, Md., who assisted in preparing the sample Computer Output Microfilm.

Dr. Daniel M. Forsyth, vice president; Russell Ham; Richard Martin and other members of the Advanced Systems Division of Information International, Inc., Los Angeles, Calif., assisted in applying the Computer Input Microfilm techniques.

CONTENTS

Single free copies of this report are available on request to:

Hydrologic Data Laboratory
Plant Physiology Institute
ARS-USDA
Room 218, Building 001
Beltsville, Md. 20705

COMPUTER INPUT MICROFILM (CIM) FEASIBILITY STUDY

By J. B. Burford and J. M. Clark[1/]

SUMMARY

This feasibility study determined that Computer Input Microfilm (CIM) techniques can be used with a high degree of accuracy to convert hydrologic data recorded in Computer Output Microfilm (COM) to computer-readable magnetic tape. Many beneficial aspects that will improve overall accuracy and reduce conversion costs were determined also.

After the proposed changes have been incorporated into the COM procedures, and as funds and time permit, a confirmation CIM production run is planned.

Computer Output Microfilm will be used as the backup medium for ARS hydrologic data-bank storage. Data format will be compatible with that required for CIM conversion to magnetic tape.

INTRODUCTION

The Agricultural Research Service's Hydrologic Data Laboratory is responsible for developing and maintaining a bank of hydrologic data and related information obtained at the various ARS Watershed Research Centers. These headquarters are located at Athens, Ga.; Boise, Idaho; Burlington, Vt.; Chickasha and Stillwater, Okla.; Columbia, Mo.; Coshocton, Ohio; Oxford, Miss.; Temple, Tex.; Tucson, Ariz.; and University Park, Pa.

Computer-readable magnetic tapes are used as the medium for manipulating and storing the active volumes of data. The value and uniqueness of the data dictate that a dependable and positive backup system be used to insure against unforeseeable mishaps and disasters such as total loss of data.[2/] Since information stored as magnetic charges in a metallic oxide coating of plastic film is quite vulnerable to such threats as stray magnetic fields, uncontrolled environment, and deterioration of the plastic film during extended periods, special facilities that have a controlled environment are required (fig. 1).

1/ Hydraulic engineer and computer specialist, respectively, Hydrologic Data Laboratory, Plant Physiology Institute, Northeastern Region, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Md. 20705.

2/ Panorama. Business Systems Market Div., Eastman Kodak Co., March 1973.

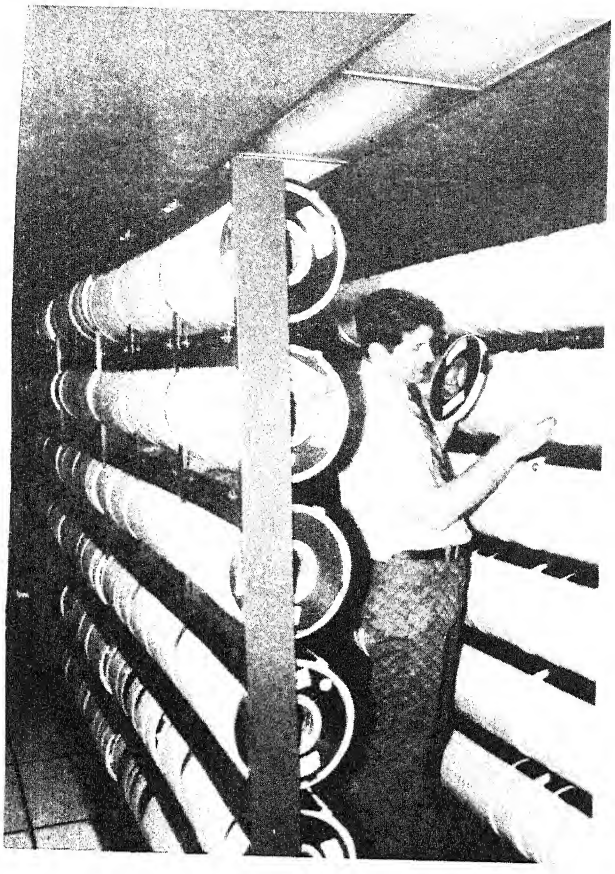

Figure 1.--Special facilities are required to store the hydraulic data bank recorded in computer-acceptable magnetic tape.

Acceptable integrity of the information stored on magnetic tape is difficult to maintain; therefore, a system of purging the tapes at some regular interval must be used, which requires computer time and expense. The Hydrologic Data Laboratory developed techniques using 16-mm Computer Output Microfilm, rather than magnetic tape, as the medium for the backup copies of the data volumes. Procedures were also studied for converting the COM-recorded data back to magnetic tape using CIM techniques. A study has been conducted by this Laboratory to determine the feasibility of these techniques. In contrast with magnetic tape, the microfilm is electronically stable, easy to store, more compact, more economical, and "human readable," besides being machine readable (fig. 2).

Information in narrative and graphic form that supplements the digitized hydrologic data is a significant part of the data bank. This information is recorded in 16-mm microfilm at the Hydrologic Data Laboratory (fig. 3) and stored with the microfilmed (COM) data.

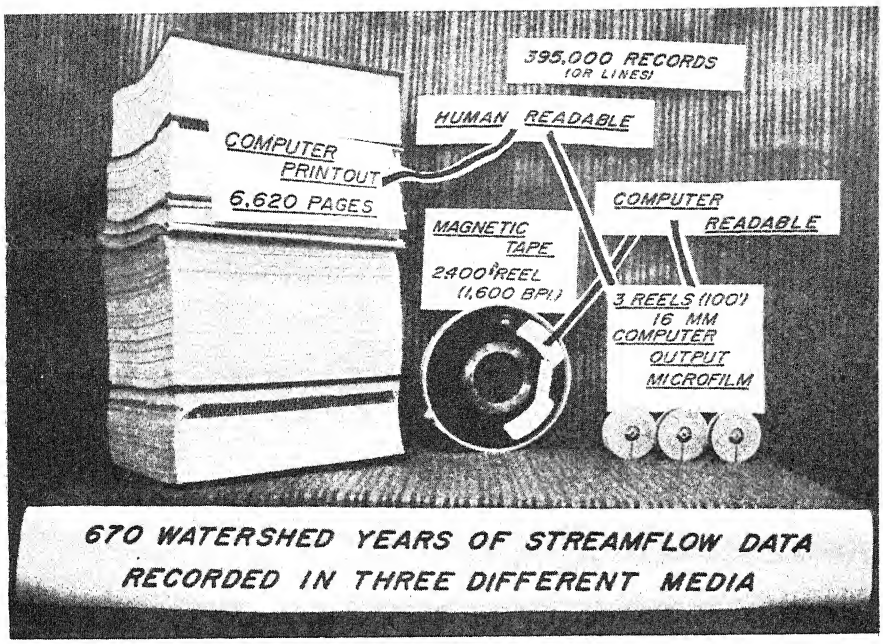

Figure 2.--A visual comparison of information stored on three different media that emphasizes some advantages of Computer Output Microfilm.

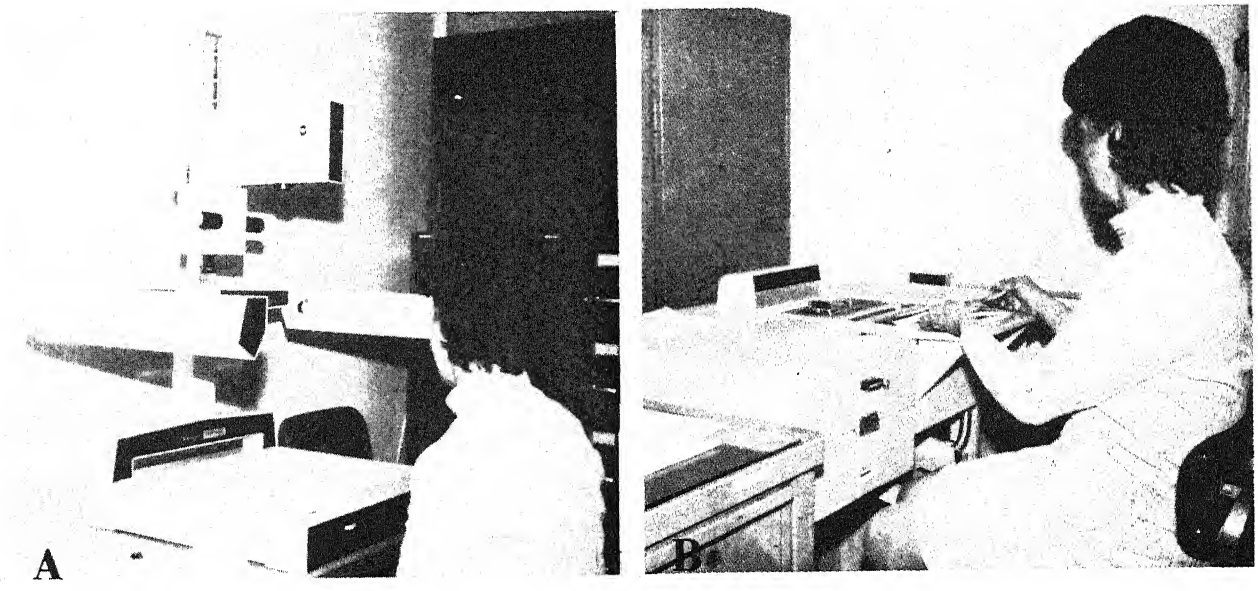

Figure 3.--Microfilming equipment used by the Hydrologic Data Laboratory to record supplemental information on 16-mm microfilm.
(A) Planetary camera used to photograph odd-sized documents.
(B) Rotary camera used to photograph documents of regular, uniform size.

DISCUSSION

Through cooperative efforts, CIM techniques that use a combination of the cathode-ray tube and optical character recognition principles were developed to read the COM backup copies and to convert them to magnetic tape, if required. This is a brief report of a CIM feasibility study that was recently completed by the Hydrologic Data Laboratory.

The Laboratory obtained COM copies of sample hydrologic data stored on magnetic tape to use in the study. The COM copies were obtained with a service-bureau-operated FR-80 COM recorder (fig. 4). The COM data were in the standard, human-readable, hard-copy printout format, reduced (24X) as required to fit on 16-mm unsprocketed microfilm. A review of the sample hydrologic COM data indicated that the image-processing system that had been developed by Information International, Inc., in the GRAFIX I, was capable of converting the microfilm images back to the computer-readable, magnetic-tape form.3/

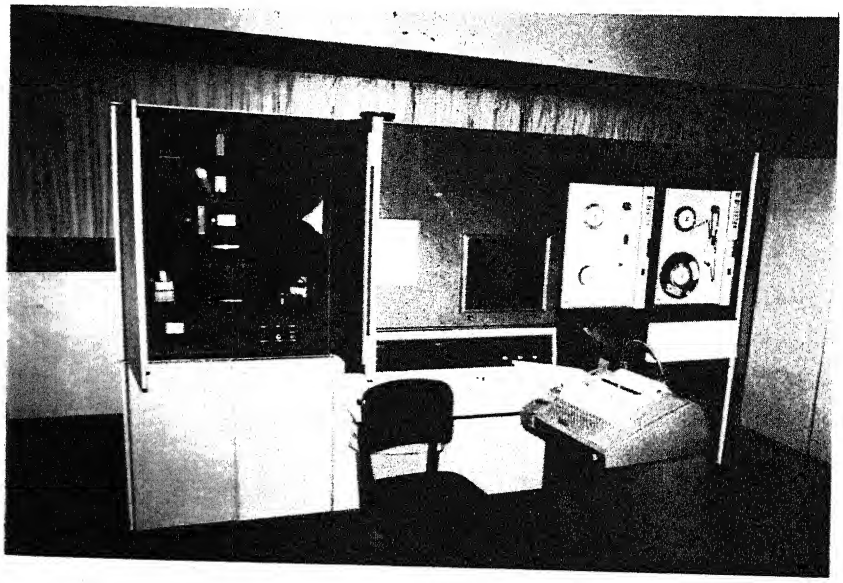

Figure 4.--Computer Output Microfilm Recorder. Dual magnetic-tape drives on the right and a microfilm camera seen through open doors on the left are controlled by the minicomputer and console in the center.

3/ Gray, S. B. Technical Description of the GRAFIX I Image Processing System. Information International, Inc. 15 pp. Los Angeles, Calif. 90340. 1971.

A contract was negotiated to convert about 265,000 characters from COM to a magnetic-tape (CIM) version, with an accuracy goal of 99.5 percent, and to furnish the Hydrologic Data Laboratory with a copy of the generated tape for a character-by-character comparison against the original tape.

The GRAFIX I image-processing system is designed to substitute (or to reject and flag) character images that do not compare with benchmark characters within specified limits. The degree of substitution or rejection can be influenced by the limits set in the system.

The CIM-generated tape contained 2,999 records with 99 characters per record, for a total of 296,901 characters. The character-by-character comparison, using computer logic, determined that 63 characters had been rejected or misinterpreted. This represents an average of 1 rejected or wrong character for each 4,173 characters read, for an accuracy of 99.98 percent. This margin of error is well within the accuracy goal. The 63 rejected or wrong characters were distributed among 53 data records, or 1 incorrect record for each 56.5 records converted, for an accuracy of 98.22 percent.

CONCLUSIONS

Many beneficial facts were realized or confirmed from the study. These facts were:

1. Hydrologic data recorded in computer output microfilm can be converted to magnetic tape with a high degree of accuracy. This conversion was accomplished at a cost of $0.75 per 1,000 characters, excluding the cost required to modify and develop the software.

2. Difficulties are encountered in keeping the image recognition system oriented when two or more adjacent blank spaces occur in a record. Many of the 63 errors resulted from efforts to handle multiple blank-space situations.[4/]

3. Several errors that occurred resulted from difficulties encountered in recognizing the decimal (.) character.[5/]

4/ A COM data format designed to eliminate multiple blank spaces should eliminate most conversion errors by filling blank-data fields with zeros and by allowing only one blank column between data fields.

5/ A physically larger decimal font used in creating the COM would help to eliminate the problem of a decimal (.) character recognition.

4. The cost of CIM-generated, magnetic-tape data is related directly to the quantity of characters converted. The COM data format should be designed so that only the data that are required for regeneration are converted.

5. Raster marks included in each frame of COM data are very helpful in the orientation of the image recognition system. Forms overlay techniques available in the COM equipment can be used to add these fiducials.

6. A system for checking the integrity of the CIM-generated data should be built into the COM version. This system would entail an algorithm to review the several digits in each record and then to compute a single value and record it at the end of the record. This value could be checked with the same algorithm after the CIM conversion.

7. The nature of hydrologic data is such that questionable character images should be rejected and flagged, as opposed to substituted.